Implicitly Learning to Reason in First-Order Logic (Extended Abstract)

Implicitly Learning to Reason in First-Order Logic: Extended Abstract*

Vaishak Belle[1] , **Brendan Juba**[2]
[1]University of Edinburgh & Alan Turing Institute
[2]Washington University in St. Louis
vaishak@ed.ac.uk, bjuba@wustl.edu

The tension between *deduction* and *induction* is perhaps the most fundamental issue in areas such as philosophy, cognition and artificial intelligence. The deduction camp concerns itself with questions about the expressiveness of formal languages for capturing knowledge about the world, together with proof systems for reasoning from such *knowledge bases.* The learning camp attempts to generalize from examples about partial descriptions about the world. In an influential paper, Valiant (2000) recognized that the challenge of learning should be integrated with deduction. In particular, he proposed a semantics to capture the quality possessed by the output of (probably approximately correct) PAC-learning algorithms when formulated in a logic. Although weaker than classical entailment, it allows for a powerful model theoretic framework for answering queries.

From the standpoint of learning an expressive logical knowledge base and reasoning with it, most PAC results are somewhat discouraging. For example, in agnostic learning, where one does not require examples (drawn from an arbitrary distribution) to be fully consistent with learned sentences, efficient algorithms for learning conjunctions would yield an efficient algorithm for PAC-learning DNF (also over arbitrary distributions), which current evidence suggests to be intractable (Daniely and Shalev-Shwartz 2016). Thus, it is not surprising that when it comes to first-order logic (FOL), very little work tackles the problem in a general manner. This is despite the fact that FOL is widely argued to be most appropriate for representing human knowledge (e.g., (McCarthy and Hayes 1969; Levesque and Lakemeyer 2001)).

In this work, we present new results on learning to reason in FOL knowledge bases. In particular, we consider the problem of answering queries about FOL formulas based on background knowledge partially represented explicitly as other formulas, and partially represented as examples independently drawn from a fixed probability distribution. Our results are based on a surprising observation made by Juba (2013) about the advantages of eschewing the explicit construction of a hypothesis, leading to a paradigm of *implicit learnability.* In the context of a reasoning task in which we wish to check if a query formula is entailed by a knowledge base, implicit learning means deciding whether or not the query is entailed by the implicitly learned knowledge base (possibly together with an additional, explicitly given knowledge base). Not only does this approach enable a form of agnostic learning while circumventing known barriers, it also avoids the design of an often restrictive and artificial choice for representing hypotheses. (A previous result that was similar in spirit only permits constant-width clauses (Khardon and Roth 1999).) In particular, implicit learning allows such learning from partially observed examples, which is commonplace when knowledge bases and/or queries address entities and relations not observed in the data used for learning.

That work was limited to the propositional setting, however. Here, we develop a first-order logical generalization. Since reasoning in full FOL is undecidable we need to consider a fragment, but the fragment we identify and are able to learn and reason with, known as *proper*$^+$, is expressive and powerful: we use clauses of the form $\forall x, y, \ldots [e \supset c]$ where e is a DeMorgan formula over atoms of the form $x = a$ for a variable x and name a, and c is a disjunction of relational literals—clauses in the usual sense. For example, with the unary relation *Mutant* and the name *logan*, we can write a clause $\forall x[x \neq logan \supset Mutant(x)]$; or with the additional binary relation $Teammate(x, y)$, we can write a clause $\forall x, y[\neg Mutant(x) \lor \neg Teammate(x, y) \lor Mutant(y)]$ (with a trivial $e = \top$). Consider that standard databases correspond to a maximally consistent and finite set of literals: every relevant atom is known to be true and stored in the database, or known to be false, inferred by (say) negation as failure. Our fragment corresponds to a consistent but infinite set of ground clauses, not necessarily maximal (Liu and Levesque 2005; Belle 2017). Moreover, the underlying language is general in the sense that no restrictions are posed on clause length, predicate arity, and other similar technical devices seen in PAC results.

To achieve the generalization, we revisit the PAC semantics and exploit symmetries exhibited by constants in the lan-

*Full version: *Vaishak Belle and Brendan Juba, Implicitly learning to reason in first-order logic. In Advances in Neural Information Processing Systems 32, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, eds., pages 3381–3391, 2019.* http://papers.nips.cc/paper/8599-implicitly-learning-to-reason-in-first-order-logic

Vaishak Belle was supported by a Royal Society University Research Fellowship. Brendan Juba was supported by NSF award CCF-1718380.

guage. Specifically, the *grounding trick* (Belle 2017) shows that to check whether a proper$^+$ KB entails a ground formula, it suffices to check if the formula can be proved using the groundings of the KB for a set of names that includes the names appearing in the ground formula and KB, together with a set of arbitrary names of size at least the rank of the KB—i.e., the maxium number of quantified variables in any formula of the KB. The grounding trick thus gives a way to check entailment while only referring to groundings using some suitable set of names.

In particular, we are able to regard a proper$^+$ KB as being *implicitly learned* with respect to a given data set if for each example (record) in the data set, some suitable grounding of that KB is satisfied by the data: By the grounding trick, checking entailment with respect to the implicit KB can be accomplished by checking entailment using that grounding. This works even if the examples in the data set only give valuations to a finite subset of an infinite universe of atoms, as in the partial observation model standard in PAC semantics. Note that if we are only given such a finite partial valuation, clauses of a proper$^+$ KB (if the equality formula e is satisfied by an infinite number of names) cannot be verified to hold, and thus such KBs cannot be learned in the usual sense—learning in the usual, explicit sense would require the ability to distinguish clauses that are true with high probability from clauses that are false on the basis of the observed data. Thus, an implicit approach is essential to learning such knowledge bases.

For example, given $Teammate(scott, logan) = \top$ and $Teammate(jean, logan) = \top$, the clause $\forall x[x \neq logan \supset (\neg Mutant(x) \vee Teammate(x, logan))]$ has groundings with $x = logan$, $x = scott$, and $x = jean$ that may be verified to hold. Similarly, given $Teammate(ororo, logan) = \top$ and $Teammate(kurt, logan) = \top$, the groundings with $x = ororo$ and $x = kurt$ (in addition to $x = logan$ again) hold. So, for this small data set with two examples, this clause will be implicitly learned; if we additionally have an explicit KB consisting of the clauses $\forall x[x \neq logan \supset Mutant(x)]$ and $\forall x, y[\neg Mutant(x) \vee \neg Teammate(x, y)) \vee Mutant(y)]$, then together these allow us to infer $Mutant(logan)$. Implicit learning enables us to answer that $Mutant(logan)$ holds when given this explicit KB together with the earlier data set. Observe that the data alone contains no values for the *Mutant* relation, and that the explicit KB alone does not allow us to infer $Mutant(logan)$. Both pieces are necessary to draw this inference.

More generally, implicit learning in this context means that we possess an algorithm for the following task. Given data consisting of partial valuations that are portions of complete valuations, and given a background KB, we can distinguish ground clauses that are falsified on the complete valuations from ground clauses that are entailed by the union of the background KB and an implicitly learned KB. Actually, more strongly, we obtain a lower bound on how often the given ground clause is satisfied on the complete valuations, assuming that the background KB is true on those valuations. This enables us to tolerate a few counterexamples (or corruptions) in the data set if we wish, achieving a kind of agnostic learning of the implicit KB. We give a quantitative bound on how much data suffices to achieve an estimate of a given accuracy with a given confidence.

Our algorithm for reasoning with implicit learning is based on a reduction to propositional entailment via the grounding trick, as discussed above. We also consider a tractable variant of the algorithm, using the approach of Liu, Lakemeyer, and Levesque (2004). This variant, which now runs in polynomial time, provides an analogous guarantee except that it distinguishes false ground clauses from those that are entailed by the union of the background KB and an implicitly learned KB at the specified limited reasoning level.

We hope our results will renew interest in learnability for languages with quantificational power. A variety of questions are left open by our work. The most complelling may be whether and how we can extend the approach to answer queries with quantifiers. It would also be interesting to extend the approach to other languages that are richer in various respects.

References

Belle, V. 2017. Open-universe weighted model counting. In *AAAI*, 3701–3708.

Daniely, A., and Shalev-Shwartz, S. 2016. Complexity theoretic limitations on learning DNF's. In *COLT*, 815–830.

Juba, B. 2013. Implicit learning of common sense for reasoning. In *IJCAI*, 939–946.

Khardon, R., and Roth, D. 1999. Learning to reason with a restricted view. *Machine Learning* 35(2):95–116.

Levesque, H. J., and Lakemeyer, G. 2001. *The logic of knowledge bases*. Cambridge, MA, USA: The MIT Press.

Liu, Y., and Levesque, H. J. 2005. Tractable reasoning in first-order knowledge bases with disjunctive information. In *AAAI*, 639–644.

Liu, Y.; Lakemeyer, G.; and Levesque, H. J. 2004. A logic of limited belief for reasoning with disjunctive information. In *KR*, 587–597.

McCarthy, J., and Hayes, P. J. 1969. Some philosophical problems from the standpoint of artificial intelligence. In *Machine Intelligence*, 463–502.

Valiant, L. G. 2000. Robust logics. *Artificial Intelligence* 117(2):231–253.